Economic evaluation of medical tests at the early phases of development

Frempong, Samuel; Sutton, Andrew; Davenport, Clare; Barton, Pelham

DOI:
10.1080/14737167.2018.1411194

Document Version
Peer reviewed version

Citation for published version (Harvard):
Frempong, S, Sutton, A, Davenport, C & Barton, P 2018, 'Economic evaluation of medical tests at the early phases of development: a systematic review of empirical studies', *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 18, no. 1. https://doi.org/10.1080/14737167.2018.1411194

Economic evaluation of medical tests at the early phases of development: a systematic review of empirical studies

Running title: A review of early cost-effectiveness studies of tests

Samuel N Frempong, Msc[1], Andrew J Sutton, PhD[2,3], Clare Davenport, PhD[1], Pelham Barton, PhD[1]

[1] Institute of Applied Health Research, University of Birmingham, Birmingham, UK

[2] Faculty of Medicine and Health, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

[3]NHIR Diagnostic Evidence Co-operative Leeds

Address for Correspondence:

Andrew J. Sutton

Faculty of Medicine and Health,

Leeds Institute of Health Sciences,

University of Leeds

LS2 9JT, UK.

Telephone: +44(0)113 343 9814

Fax: +44 (0)113 343 5740

Email: a.j.sutton@leeds.ac.uk

WORD COUNT: 4,960

The authors did not receive any financial support and have no conflicts of interest

ABSTRACT

Background: There is little specific guidance on the implementation of cost-effectiveness modelling at the early stage of test development. The aim of this study was to review the literature in this field to examine the methodologies and tools that have been employed to date.

Methods: A systematic review was conducted to identify relevant studies in established literature databases.

Results: Five studies were identified and included for narrative synthesis. These studies revealed that there is no consistent approach in this growing field. The perspective of patients and the potential for value of information (VOI) to provide information on the value of future research is often overlooked. Test accuracy is an essential consideration, with most studies having described and included all possible test results in their analysis, and conducted extensive sensitivity analyses on important parameters. Headroom analysis was considered in some instances but at the early development stage (not the concept stage).

Expert commentary: The techniques available to modellers that can demonstrate the value of conducting further research and product development (i.e. VOI analysis, headroom analysis) should be better utilized. There is the need for concerted efforts to develop rigorous methodology in this growing field to maximize the value and quality of such analysis.

Key words: systematic review; VOI and Headroom analysis; early cost-effectiveness; economic evaluations; modelling; medical tests

1.0 INTRODUCTION

Economic evaluation is a tool that decision makers typically use to compare competing interventions after having demonstrated their quality, safety and effectiveness [1], and identifies those that are cost-effective with respect to their costs and consequences [2,3].Decision maker may refer to an established institution such as The National Institute for Health and Care Excellence (NICE) in the UK, or any individual or group of individuals who have vested authority to use economic analysis to support decision making for example in a medical technology company. Test evaluation 'lags behind' and differs from the evaluation of treatments. Medical testing affects a patient's quality of life primarily in two ways: direct effects associated with the test or the testing process (either positive/negative) and benefits arising from clinical decisions made based on the test results. It is by capturing these effects that the true value of a test can be determined. This is a difficult undertaking, but innovators are encouraged to start engaging with clinicians early in the development process in order to delineate the pathway(s) for the new test and also to gain an understanding of issues that are likely to be important to patients (referred to as "patients' perspective" in this study, and "perspective" does not refer to the classical definition as used in economic evaluation). An in-depth understanding of the care pathway is key to understanding the optimal placement and role (the chosen role of a new test implies the necessary properties it should possess) of a test on the care pathway and identifying key outcomes to be include in the evaluation.

[Please insert Figure 1 here]

Figure 1 shows the various phases on the medical technology innovation pathway and the stages at which economic analyses are conducted (i.e., early and late phases). The early phase comprises the concept stage and the early development stage. The concept stage is the discovery and ideation phase. In the case of medical tests, the test is still hypothetical at this

point with no available data on test parameters (e.g. accuracy, cost, etc.). The early development phase occurs between the end of the concept stage and the equivalent of the Phase I stage of clinical trials of drugs. It involves the assessment of certain test properties and some form of experimental data may be available at this stage.

In health care systems worldwide, economic evaluations are usually conducted at the late stage of product development [4], for example after Phase III drug trials (Figure 1). The rationale being that this is the point at which there is sufficient product-specific data for their proper evaluation [5]. In the case of medical tests, test performance would be assessed at the preclinical stages, test accuracy examined in phases 1 and 2, and clinical effectiveness assessed in phases 3 and 4 [6] (Figure 1). Coverage and reimbursement decisions are therefore made on tests at the time when substantial resources have already been committed to their research and development; thus, any negative coverage and reimbursement decisions would lead to no returns on investment and a loss being incurred (loss to manufacturers; opportunity cost for research; and inefficient use of Health Technology Assessment resources) [7,8]. Furthermore, advances in medical technology typically occur more rapidly than for drugs [9], and so leaving an economic evaluation to the late stages of development may make any new findings redundant. Thus, there is now an increasing interest in the economic analyses of medical tests at their early phases of development by investors, innovators, and policy makers, to identify their potential economic value and likely impact [10,11].

The early economic evaluation of a medical technology is defined as an iterative economic evaluation process to assess its economic value and likely impact [11], and is applied in its development process at a point where it can still be considered experimental or emerging [8]. In the case of medical tests, this is usually from the concept stage up to stage I clinical trials [12] and provides useful information that informs investment and design decisions under

conditions of high uncertainties before the clinical performance of the test is established [4]. A key characteristic of early economic evaluation is the use of limited data which is associated with increased uncertainty and is likely to be more pertinent for tests [13,14]. If there is a considerable amount of uncertainty associated with data, as will be the case for the early economic evaluation, value of information (VOI) analysis is essential [15]. This provides an analytic framework for decision makers to decide whether evidence is sufficiently robust to recommend investing into the further development of a test or not, and if not, identify specific areas where further information is needed to decide on further development decisions [16]. This will enhance the efficient use of limited resources and potentially reduce the risk of investing in a test which is not economically viable. Furthermore, the ubiquitous existence of uncertainty in parameter estimates and model structure at the early stage warrants the need to include extensive sensitivity analysis to test the robustness of parameter estimates, determine the range of parameters which have the greatest impact on cost-effectiveness, and determine how sensitive the results are to changes in model structure [17]. Another key issue is the "maximum cost" at which the index testing strategy is still cost-effective at a given willingness-to-pay threshold. The headroom approach is a simple but very useful method that provides the framework for estimating this value (known as the headroom) [18]. Although the analysis is done at the early development stage, the rationale is that eventually when the test is fully developed, its cost-effectiveness will be assessed at a giving willingness-to-pay threshold (for example £20-30,000) in the UK, thus the headroom is estimated at that threshold. Estimated headroom provides valuable information to support decision making on the feasibility of further test development and its consideration at the early stages especially at the concept stage will promote efficiency (i.e., if estimated headroom is too low, and realistically it is not possible to develop the new test below this level, resources should not be committed to its further development).

Early economic evaluation of medical tests has potentially profound advantages for both decision makers and innovators. For decision makers, the early identification of the economic value and likely impact of new tests could help allocate limited budgets more efficiently by identifying which tests to fund for further development and which tests to reimburse on condition of further data collection (known as “coverage with Evidence Development” as practiced in the USA and UK). Furthermore, early economic analyses studies speed up decision making regarding test adoption in the late phase of development and support the management of test diffusion through “horizon scanning”: the early identification of new economically viable medical devices [8]. Early economic evaluation is becoming increasingly important as there is a growing demand to demonstrate value for money; however, there is little specific guidance on their implementation [11]. Early evaluations are more iterative in nature and conducted at a time where there is much less available data compared to the late phase of test development [10]. This is compounded by the fact that tests are indirect in terms of their impact on patients, therefore data on accuracy and downstream consequences may only be obtained from mapping clinical pathways resulting in potentially more uncertain data. These differences suggest that the methods used in the analysis of late economic evaluation need modification for use in early evaluation. The question therefore arises, how have economic evaluations conducted during the early phases of test development been done to date, and can any lessons be learnt from them? To identify the current practice in this field requires an up to date and focussed review of previous early economic evaluations of medical tests.

The aim of this systematic review was to examine the methodologies and tools that have been employed in early economic evaluation studies of medical tests, specifically to:

1) Gain a greater understanding of how the problem of insufficient data for model parameterisation has been resolved.

2) Understand whether and how testing pathways have been modelled.

3) Examine whether sensitivity analysis has acknowledged the uncertainty that accompanies early modelling and the stage at which it has been undertaken.

2.0 METHODS

The following databases were searched for any studies published from inception to July, 2016.

I. Cochrane Library (CENTRAL)
II. Medline
III. EconLit
IV. Centre for Reviews and Dissemination [Database of Abstracts of Reviews of Effects (DARE), Heath Technology Assessment (HTA), and NHS Economic Evaluation Database (NHS EED)]
V. Health Management Information Consortium (HMIC)
VI. Excerpta Medica dataBASE (EMBASE)

Internet searches were also conducted (e.g., Google scholar and the websites of organisations related to innovations in health care such as EuroScan) to identify any grey literature. The nature of early economic evaluations means that it is unlikely to be published. Therefore, medical technology companies (7) were contacted to identify unpublished evaluations to complement the literature review, but all efforts proved futile. The reference list of the relevant articles included in the review were scanned for additional relevant articles. The list of articles used was managed through the reference management software, Endnote.

2.1 Search terms

The search strategy was customized for each database and used a combination of MESH terms and index terms as shown in Table 1. The search strategy was developed in

consultation with an information specialist and piloted to ensure that all relevant studies were retrieved. The complete search strategies including how the different question elements were combined are illustrated in Table 2.

[Please insert Table 1 here]

[Please insert Table 2 here]

2.2 Inclusion criteria

Primary studies were deemed potentially relevant to be included in the review if they were:

1. An economic evaluation conducted at the early phases of a medical test development.
2. The evaluated technology was a medical test or series of tests used together (at least one test needed to be present in at least one arm of the analysis).

2.3 Exclusion criteria

Articles were excluded from the review if they were:

1. Trial protocols or commentaries.
2. Letters or editorials.

2.4 Selection of articles for the review

After the removal of duplicates, a 2-stage screening of titles and abstracts followed by full text articles was undertaken against the inclusion criteria by two reviewers (S.F. and P.B.) independently. Disagreements regarding study eligibility were resolved using the opinion of a third reviewer (C.D.) where necessary. All studies identified after the second stage of article selection were subsequently considered for data extraction.

2.5 Data extraction

Data extraction was conducted by (S.F. and A.S.) for each study included in this review to answer the following questions:

- How was the problem of an early analysis not having sufficient data for model parameterisation resolved?
  - What sources (type) of data were used?
  - Was the source (type) of data used influenced by the stage of evaluation?
- Modelling of testing pathways
  - What type of model was used in modelling disease progression and the testing pathways (e.g. Decision tree, Markov)?
  - Did the studies consider test accuracy and all possible test results (i.e. true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP))?
  - Did the analysis include all the issues important to patients on the test-treat pathway (e.g., personal costs incurred when accessing testing, effect of testing pathway on quality of life)?
  - Was headroom analysis included, and at what stage of the analysis?
- Uncertainty
  - Did sensitivity analysis acknowledge the uncertainty that accompanies early modelling?
  - Was value of information (VOI) analysis conducted?

2.6 Quality assessment

The methodological quality of the included studies was assessed using a 10-point checklist for economic evaluations [19], and a score was assigned based on how well criteria were met; scores of 1, 0.5 and 0 were assigned to “yes”, “cannot tell” and “no” respectively. Thus, each study scored from 0 (worst) to 10 (best) [20]. However, because the focus of this systematic review was to explore methodologies used and not to comment on the validity of results and conclusions drawn from these studies, no study was rejected on quality grounds. It is worth mentioning that one of the main issues when scoring the quality of publications is how to

weight each item to provide an overall quality score. The quality issues raised by different clinical topics differs. There is no objective way of doing this, thus, one has to be cautious since items are not always equivalent. Different methods are likely to produce different scores and in some cases scoring might even bias quality assessment. However we do not think the potential limitations of quality scores are applicable to this study as assessment of quality was concerned with the presence or absence of a methodological approach and not the effect of an approach on outcome measure.

3.0 RESULTS

After de-duplication, 4,494 unique articles were identified for title and abstract screening. 88 titles and abstracts were potentially eligible for inclusion. After full text screening, and unsuccessful attempts to contact medical technology companies, manufacturers of the technologies identified from the EuroScan website and obtain full text of conference abstracts, five studies were included for narrative synthesis. The PRISMA diagram (Figure 2) illustrates the results of the screening process with reasons for exclusion noted.

[Please insert Figure 2 here]

[Please insert Table 3 here]

3.1 Characteristics of included studies

Table 3 illustrates the characteristics of the studies included in this review which were published between 2005 and 2016 [21-25]. Two studies were conducted in the Netherlands [23,25] and three studies were conducted in the USA [21,22,24]. The clinical conditions explored spanned a range of disease areas, namely rheumatoid arthritis, persistent asthma, chlamydia trachomatis, peripheral artery disease, and coronary artery disease. All studies stated the intended applications of the tests under consideration: these being diagnosis, predicting a response to treatment, screening, and risk assessment. Four studies were

conducted at the early development stage (Figure 1) [22-25] and one study was conducted at the concept stage [21]. Model-based approaches were used in cost-utility [21,23,24,25] and cost-effectiveness analyses [22]. Four studies were conducted from the societal perspective [21,23,24,25] and one study was conducted from the public healthcare perspective [22]. Three studies adopted a time horizon of ≤ 10yrs [22,24,25], one study adopted a lifetime time horizon [23], and one study did not incorporate a time horizon [21]

3.2 Quality assessment of included studies

Based on the quality assessment criteria applied, two studies had a score of 8.5 [22,25], two studies had a score of 8 [23,24] and one study had a score of 7 [21] (Table 4). All studies lost a point each for not giving a comprehensive description of the testing strategies and therefore provided insufficient information about the clinical pathway, making it difficult to tell whether any important alternatives were omitted from the studies [21-25].Three studies lost a point each for not identifying and including all important outcomes for each alternative (such as those from the patient's perspective) [21,23,24]. One study lost a point because it was not clear whether cost and consequences were adjusted for differential timing. And whether they were valued credibly because all the data used were based on assumptions but no information on the basis of the assumptions was provided [21]. It is worth mentioning that due to word count restriction not all the clinical information can be incorporated in economic evaluations. Usually and especial for piggy-bag evaluations, these have to be read together with clinical publications to get an impression of the total available evidence on a topic or research question for evaluation. In this study, this was done before and during quality assessment to ensure an effective assessment.

[Please insert Table 4 here]

3.3 Data extraction

3.3.1 How was the problem of an early analysis not having sufficient data for model parameterisation resolved?

To populate the index testing strategy arm of the models, several sources of data were used across the different studies from four main perspectives (test accuracy, costs, measures of effectiveness and transitional probabilities describing the disease states) as shown in Table 5, and these were found to be influenced by the stage of evaluation/analysis.

[Please insert Table 5 here]

One study was conducted at the concept stage [21] and all data used were based on assumptions. Interestingly, no information on what these assumptions were based was provided. Sources of data describing the test accuracy estimates, costs, measures of effectiveness and transitional probabilities varied across studies conducted in the early development stage (Table 5). It is noted that, in all these studies, the plausibility of the estimates and the robustness of the results obtained by employing these estimates were examined in sensitivity analysis.

3.3.2 Modelling of testing pathway

3.3.2.1 What type of model was used in modelling disease progression and the testing pathways?

Three studies used a Markov state transition model [21,23,24], one study used a decision tree [22] and one study used both the decision tree and Markov model [25]. The Markov models covered the disease and the decision trees covered the testing pathways. Information was provided to justify the model structure in only three studies [23-25].

3.3.2.2 Did the studies consider test accuracy and all possible test results?

Four studies modelled explicitly each of the possible test results [21,22,24,25]. One study only considered TP and TN, and assumed a perfect biomarker; however, no information was

given on what the consequences of test errors would be for this test [23]. In one study, the first year of the 5-year time horizon was modelled as a decision tree with chance nodes at 6 and 12 months (as repeated testing is part of the clinical pathway) to classify patients as TP, FP, TN and FN with those classified as TP or FN at 12 months entering the patient level state transition model and followed for 4 years [25]. In another study, four subpopulations based on the test result of TP, FN, TN and FP were considered within the same model [24].

3.3.2.3 Did the analysis include all the issues important to patients on the test-treat pathway?

It is notable that of the four studies that were conducted from the societal perspective, only one study acknowledged and included some issue on the test-treat pathway that may be relevant to the patient [25]. In the other studies, the issue of the patient perspective was ignored [21,23,24]. In Buisman et al [25], follow-up visit costs and productivity costs (which was defined as the number of days that a patient with a paid job was absent from work) were included. The study by Huang et al [22] considered the effects of how long patients were willing to wait to obtain their test results. This seems reasonable as the study was conducted from a public health care perspective and one of the key mechanisms by which the test in this study might impact on outcomes is being able to treat patients at the time they present to prevent onward transmission of infection. Indeed, in this study, one-way sensitivity analyses demonstrated that one of the key parameters driving the results of the cost-effectiveness analysis was how long patients were willing to wait to obtain their results. This was because a short processing time reduced the time between testing and treatment thereby increasing treatment rates and subsequent improvement in the quality of life of patients at a population level.

3.3.2.4 Was headroom analysis included, and at what stage of the analysis?

Headroom analysis was included in the two studies conducted in the Netherlands [23,25]. It is noted that neither of these studies was conducted at the concept stage of development but at the early development stage when there was some data available to describe the test parameters. Thus, headroom analysis was included at the early development stage but not the concept stage.

3.3.3 Uncertainty

3.3.3.1 Did sensitivity analysis acknowledge the uncertainty that accompanies early modelling?

In all studies, the issue of a lack of data and the simplification of models to represent reality were acknowledged as study limitations. To deal with these limitations, extensive sensitivity analysis was undertaken on all important parameters (e.g., sensitivity analysis of test accuracy) to evaluate the influence of uncertainty on model predictions. Probabilistic and deterministic sensitivity analyses were conducted in four studies [22-25] and deterministic sensitivity analysis was conducted in one study [21].

3.3.3.2 Was value of information (VOI) analysis conducted?

Considering the importance and role of VOI analysis in supporting decision making under conditions of uncertainty (typically characteristic of early economic evaluation), it is notable that this was not conducted in any of the included studies.

4.0 DISCUSSION

Typically, economic evaluations are undertaken as a one-off exercise at the late stage of development of a new medical device [26]. Increasingly, several studies have indicated the importance of an iterative use of economic evaluation during the early phases of development of medical devices to identify their potential economic value and likely impact and to support

and guide decision making under conditions of high uncertainty [27,28]. However, there is little specific guidance on their implementation to maximize the value and quality of such analysis. This systematic review focussed on exploring the approaches used in early economic evaluation of medical tests. Five studies were identified and data was extracted from these studies to gain insight into how the problem of an early analysis not having sufficient data for model parameterisation has been resolved, understand whether and how testing pathways have been modelled, and examine whether sensitivity acknowledged the uncertainty that accompanies early modelling.

The major issue associated with test evaluation is the rather difficult connection between the test and final health outcomes. The evaluation of tests has typically been restricted to test accuracy (sensitivity and specificity) an intermediate outcome measure; which may influence but do not directly determine patient relevant outcomes. However, for economic evaluation of tests, we need to know the longer term costs and effects of the test-treat pathway: patient outcomes dependent on test results. This usually involves extensive modelling of delineated test-treat pathways. This is difficult enough to do for an established test, but for a new test it is even more complex and difficult because pathways may not be defined. This may be a plausible reason why there are so few publications on early test evaluation and thus the small number of studies identified and included in this study.

To resolve the issue of an early analysis not having sufficient data for model parameterisation, studies in this review relied on different sources of data and the source of data used was influenced by the stage at which the analysis was conducted (concept stage or early development stage). One study was conducted at the concept stage and, as expected, the relevant data on test accuracy, costs, measures of effectiveness and transitional probabilities of the index testing strategy were all based on assumptions. However, the basis for these assumptions was not stated, making these estimates somewhat arbitrary and weakly

informed, although their plausibility and robustness were extensively examined in sensitivity analyses. Four studies were conducted at the early development stage, yet, within the confines of this stage of evaluation, diverse sources of data were used to inform the new testing strategies across the different studies. This observation is explained by the fact that the tests were at various stages of development even within the early development stage; hence, different levels of data were available specific to different tests. For example, in one study, the test was in an experimental phase and primary data were available for most parameters for the analysis. In other studies, primary data were not available for all parameters; hence other plausible sources of data including expert opinion and secondary sources were relied upon to supplement primary data (making investigation of uncertainty particularly important). The plausibility and robustness of the estimates were however examined in sensitivity analyses and their effects on the conclusions drawn from the models examined.

Test accuracy was considered in all five studies and four studies modelled explicitly each of the possible test results. This meant that the full implications of test accuracy on the model results could be examined in the analysis. In the study that did not model each of the test results, this meant that it was not possible to determine the consequences of test errors. However, it is notable that even though a societal perspective was adopted in four studies, only one of these four considered aspects of the test-treat pathway that may be important to patients. This is important because patient perspectives can have a significant impact on the conclusions drawn from a model. For example, in one study, one of the key parameters driving the results of the cost-effectiveness analysis was how long patients were willing to wait to obtain their test results. Overlooking this important parameter in this particular analysis would have led to misleading conclusions. The issue of the timing of test results (rapid tests) is probably the most obvious (and most studied) way in which a test might produce benefits aside from improvements in accuracy [22]. However, it is important to

stress that other aspects such as test acceptability to patients and professionals, procedural harms or benefits of the testing process etc., should also be considered [29,30].

Headroom analysis was included in two studies, and it was noted that neither of them was conducted at the concept stage of development. However, in the headroom paradigm, assumptions can be made at the concept stage for a preliminary assessment of whether a test warrants further development first. Later, in the early development stage when more evidence becomes available, this can then be updated to determine the headroom in the face of newly available evidence. This reduces the risk of investing in a bad technology. Though it might be argued that if a project is terminated at the early development stage, not many resources would have been invested, it is equally true that, if the headroom had been established at the concept stage, the invested resources could have still been used efficiently elsewhere. Headroom analysis has potential in the early economic evaluation of tests to promote efficiency at the beginning of the test development process.

Sensitivity analysis is important in early economic evaluations where there is not enough device-specific data, and initial model parameter estimates may have to be derived from data sources associated with high levels of uncertainty. The emphasis here is whether sensitivity analysis acknowledges the uncertainty associated with early evaluation and the early stage at which it is conducted. All the studies reviewed acknowledged the uncertainty associated with data used in their analysis: extensive sensitivity analyses were performed on important parameters to evaluate the influence of uncertainty on model predictions. However, it is notable that VOI analysis was not conducted in any of the included studies, meaning that no insights into the value of future research were obtained. Thus there was the possibility of drawing incorrect conclusions from the results about whether to further fund the development of a new testing strategy or not based on the evidence used. This could have been because the

studies were conducted at a time when VOI was not well established as a concept or it was deemed irrelevant.

4.1 Limitations of the research

Few studies were found to be eligible to be included in the review. Furthermore, studies were conducted to answer different research questions, so it was not possible to compare their results. There is no existing quality assessment tool specifically to evaluate an economic evaluation focussed on testing (which requires development of a novel tool).

4.2 Recommendations

To improve future practice, and based on the study findings, the following recommendations for early economic evaluations focussed on testing conducted in the future are made:

- To fully capture the potential effects of testing on patient relevant outcomes and thus the potential health economic impact of tests, it is crucial that the assessment of the outcomes of testing goes beyond health and specific payer perspectives to acknowledge and include issues on the testing pathway that may be relevant to patients (including anxiety, acceptability, and loss of income). This will mitigate against over- or underestimating the true value of tests in early modelling studies and thereby appropriately inform decision making.
- All possible test results and their subsequent patient pathways should be described in a model to ensure that the full implications of test accuracy are considered in the analysis.
- Extensive sensitivity analyses of all important parameters should be undertaken to evaluate the influence of uncertainty on model predictions.
- The potential adoption of VOI should be considered. This can be beneficial in mitigating against drawing wrong conclusions from study results to fund the further development of a new testing strategy based on the available evidence used.

- Headroom analysis has potential to provide important insights into the viability of developing new tests and should be considered at the early stages of test development, especially at the concept stage to promote efficiency at the start of the test development process.

4.3 Suggestions for future work

The review has shown that some of the methods proposed (VOI and Headroom analysis) are not being used. Is there a need to refine/develop these methods for the specific context of early economic evaluation of medical tests? Or are other tools needed? To answer these questions, further research is needed in this field. Currently, there is no existing quality assessment tool specifically to evaluate an economic evaluation focussed on testing. Medical test evaluation is complex (tests are indirect in terms of their impact on patients) and differs from the evaluation of treatments. Thus, there is the need to develop a tool to capture such complexities if the methodological quality of economic evaluations focussed on testing is to be properly assessed. For example, the development of a checklist could prompt consideration of the ways in which tests might impact on patients (aside from accuracy).

ACKNOWLEGEMENTS

We would like to acknowledge Mrs Susan Bayliss for her insightful and constructive suggestions in developing the search strategy.

Expert commentary

In an increasingly resource-constrained environment, early economic evaluation of medical tests has potentially profound advantages for both decision makers and innovators as the

demand to demonstrate value for money increases. The early assessment of the potential economic value and likely impact of a test enhances more informed decision making that could potentially guarantee successful implementation in the future. However, there is little specific guidance on their implementation and there is no consistent approach on the methodologies and tools to be used. For early economic evaluation to become a practical tool there is the need for concerted efforts to develop rigorous methodology in this growing field to maximize the value and quality of such analysis.

Five-year view

The growing demand to demonstrate value for money is likely to be associated with an increase in the use of economic evaluation at the early phases of medical test development. With the limited guidance available on their implementation, it is expected that the inconsistent use of methods in the early economic evaluation of medical tests will continue. However, with the publication of additional studies highlighting the need for the development of rigorous methodology in this growing field, there will be an increase in awareness among researchers. We expect that this will lead to an increase in the effort to develop rigorous methodology to maximize the value and quality of such analysis.

Key issues

- There is an increasing interest in the adoption of early cost-effectiveness modelling for test evaluation. However, there is little specific guidance on their implementation.
- The results revealed that there is heterogeneity in the approaches used in this growing field. The perspective of patients and the potential for value of information (VOI) to provide information on the value of future research is often overlooked.

- Test accuracy is an essential consideration, with most studies having described and included all possible test results in their analysis, and conducted extensive sensitivity analyses on important parameters.
- Headroom analysis was considered in some instances but at the early development stage (not the concept stage).
- There is the need for concerted efforts to develop rigorous methodology in this growing field to maximize the value and quality of such analysis.

REFERENCES

1. Gazelle GS, McMahon PM, Siebert U, Beinfeld MT. Cost-effectiveness Analysis in the Assessment of Diagnostic Imaging Technologies 1. Radiology. 2005; 235(2):361-370.
2. Drummond MF, Sculpher MJ, Torrance GW, O'Brien, BJ, Stoddart, GL. Methods for the Economic Evaluation of Health Care Programmes, Third Edition, Oxford University Press, New York; 2005.

3. Jackson, D. HEALTHCARE ECONOMICS MADE EASY, First Edition, Scion Publishing Ltd, Oxfordshire; 2012.
4. Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. International Journal of Technology Assessment in Health Care. 2008; 24(01):36-44.
5. Redekop K, Mikudina B. Early medical technology assessments of medical devices and tests. Journal of Health Policy and Outcomes Research. 2013; 1: 23-37.
6. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Medical decision making. 1991; 11:88-94.
7. Hartz S, John J. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. International Journal of Technology Assessment in Health Care. 2008; 24(04):465-472.
8. Hartz S, John J. Public health policy decisions on medical innovations: What role can early economic evaluation play?. Health Policy. 2009; 28;89(2):184-192.
9. NICE. Diagnostics Assessment Programme manual, 2011. Available from http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0ahUKEwirjvTuyurKAhVJQBQKHe77CS8QFggcMAA&url=http%3A%2F%2Fwww.nice.org.uk%2FMedia%2FDefault%2FAbout%2Fwhat-we-do%2FNICE-guidance%2FNICE-diagnostics-guidance%2FDiagnostics-assessment-programme-manual.pdf&usg=AFQjCNEi6_ApSXVuG2cah8Oz1oYXNVh5nA&bvm=bv.113943665,d.bGs [Accessed on 23/09/2015].
10. Tu HA, Han D, Pechlivanoglou P, Miller F, Rac V, Chin W, Tricco AC, Paulden M, Bielecki J, Krahn M. Early economic evaluation of emerging health technologies: protocol of a systematic review. Systematic reviews. 2014; 23;3(1):81.
11. Buisman LR, Rutten-van Mölken MP, Postmus D, Luime JJ, Uyl-de Groot CA, Redekop WK. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. International journal of technology assessment in health care. 2016; 32(1-2):46-53.

12. Markiewicz K, van Til JA, IJzerman MJ. Medical devices early assessment methods: systematic literature review. International journal of technology assessment in health care. 2014;30(02):137-146.
13. Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis. Health Technology Assessment. 2004; 8(27).
14. Tuffaha HW, Gordon LG, Scuffham PA. Value of information analysis in healthcare: a review of principles and applications. Journal of medical economics. 2014;17(6):377-83.
15. Briggs A., Sculpher M., Claxton K. Decision modelling for health economic evaluation, Oxford University Press, New York.; 2006.
16. Chen MH, Willan AR. Value of information methods for assessing a new diagnostic test. Statistics in medicine. 2014; 33(11):1801-1815.
17. Morris S, Devlin N, Parkin D, Spencer A. Economic Analysis in Health Care, Second Edition, John Wiley and Sons, West Sussex; 2012.
18. McAteer H, Cosh E, Freeman G, Pandit A, Wood P, Lilford R. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. Journal of tissue engineering and regenerative medicine. 2007; 1(5):343-349.
19. Drummons MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. British Medical Journal. 1996; 313(7052):275-284.
20. Gonzalez-Perez JG. Developing a scoring system to quality assess economic evaluations. The European Journal of Health Economics. 2002; 3(2):131-136.
21. Bosch JL, Beinfeld MT, Muller JE, Brady T, Gazelle GS. A Cost–Effectiveness Analysis of a Hypothetical Catheter-Based Strategy for the Detection and Treatment of Vulnerable Coronary Plaques with Drug-Eluting Stents. Journal of interventional cardiology. 2005; 18(5):339-349.

22. Huang W, Gaydos CA, Barnes MR, Jett-Goheen M, Blake DR. Comparative effectiveness of a rapid point-of-care test for detection of Chlamydia trachomatis among women in a clinical setting. Sexually transmitted infections. 2013; 89(2):108-114.
23. Vaidya A, Joore MA, ten Cate-Hoek AJ, Cate HT, Severens JL. Cost–effectiveness of risk assessment and tailored treatment for peripheral arterial disease patients. Biomarkers in medicine. 2014; 8(8):989-999.
24. Wu AC, Gay C, Rett MD, Stout N, Weiss ST, Fuhlbrigge AL. Pharmacogenomic test that predicts response to inhaled corticosteroids in adults with asthma likely to be cost-saving. Pharmacogenomics. 2015; 16(6):591-600.
25. Buisman LR, Luime JJ, Oppe M, Hazes JM, Rutten-van Mölken MP. A five-year model to assess the early cost-effectiveness of new diagnostic tests in the early diagnosis of rheumatoid arthritis. Arthritis research & therapy. 2016; 18(1):135.
26. Vallejo-Torres L, Steuten LM, Buxton MJ, Girling AJ, Lilford RJ, Young T. Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach.. International journal of technology assessment in health care. 2008;24(4):459-464.
27. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. Journal of Health Services Research. 1997; 2(1):26-30.
28. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An iterative Bayesian approach to health technology assessment: application to a policy of preoperative optimization for patients undergoing major elective surgery. Medical Decision Making. 2006;26(5):480-496.
29. Ferrante di Ruffano LF, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. British Medical Journal. 2012;344:e686.
30. Ferrante di Ruffano LF, Dinnes J, Taylor-Phillips S, Davenport C, Hyde C, Deeks JJ. Research waste in diagnostic trials: a methods review evaluating the reporting of test-treatment interventions. BMC medical research methodology. 2017; 17(1):32.